DOCUMENT RESUME

ED 186 378 SP 015 837

AUTHOR Hosford, Philip L.; Martin, Jeanette V.
TITLE Historical Analysis of Videotape Usage in Predicting Teacher Effectiveness. A Presentation in the College of Education Dialogue Series.
INSTITUTION New Mexico State Univ., Las Cruces.
PUB DATE 19 Feb 80
NOTE 27p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Behavior Patterns; Educational Research; *Evaluation Methods; Interaction; *Observation; Student Teacher Relationship; Teacher Behavior; *Teacher Effectiveness; *Teacher Evaluation; *Videotape Recordings

ABSTRACT

A review of past methods of observing teacher behavior is presented, with comment on the relative effectiveness of each method provided. A research model using videotape recording of teachers' in-classroom behavior is suggested, and information given concerning the current use of similar methods of observation. (LH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186378

A DIALOGUE SERIES Monograph

Teacher Evaluation Via Video Tapes

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

by Philip Hosford

Jeanette Martin

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Philip L. Hosford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

A Presentation in the College of Education

DIALOGUE SERIES

Tuesday, February 19, 1980

SP 015 837

HISTORICAL ANALYSIS OF VIDEOTAPE USAGE IN PREDICTING TEACHER EFFECTIVENESS

A Presentation in the College of Education

DIALOGUE SERIES

New Mexico State University
Las Cruces, New Mexico

by

Philip L. Hosford

and

Jeanette V. Martin

College of Education
New Mexico State University

February 19, 1980

HISTORICAL ANALYSIS OF VIDEOTAPE USAGE IN PREDICTING TEACHER EFFECTIVENESS

INTRODUCTION

What attributes characterize a good teacher? How does one evaluate the performance of a teacher to determine whether that teacher is, indeed, effective or ineffective? These questions and similar ones have been common denominators in the research on teaching and teacher effectiveness.

Several approaches to research on teaching have been developed. Models have been formulated to explain, or define, effective teaching. A pattern of increasingly sophisticated research appears to have evolved since the 1950's. The purpose of this paper is to: (1) provide a general literature review of this developing pattern; (2) to propose a research model and describe the research related to the first section of that model utilizing videotapes; and (3) discuss testing needed of the other portions of the proposed model.

LITERATURE REVIEW OF RESEARCH ON TEACHER EFFECTIVENESS

The problem of measuring teacher effectiveness has been approached extensively since the late 1950's and early 1960's. Prior to this time much of the research done in the area was in terms of judgments, or ratings, which described the characteristics of "good" and "poor" teachers (Soar, 1972). Typical of this type of research was that of M. I. Burkhard (1962) and David Ryans (1960). These studies described certain traits, and/or attitudes which appeared to be attributable to "good" and "poor" teachers. Conducted in the hope of finding generalizations which would be applicable

to describing the effective teacher, much of this research was predicated on a relatively high level of inference, that is, it was difficult to ascertain exactly which behaviors determined such terms as "warm," "enthusiastic," "cheerful," etc. Nevertheless, these studies are significant and heralded the beginning of an era of research which by its very nature is fraught with difficulties. The classroom environment, the teacher, and the pupils all represent a complex set of relationships.

A new approach developed during the 1960's concentrated on identifying "low-inference" behaviors which were typically recorded as they occurred. This model was based upon quantitative objective measurements of specific behaviors such as verbal interactions and direct vs. indirect teaching. Correlations were sought relating specific teaching behaviors and differences in pupil achievement or attitude.

A proliferation of descriptive studies of classroom behaviors and verbal interactions followed. Observation systems of various types were developed. In the cognitive area Gallagher and Aschner (1963) devised a system classifying teacher-pupil verbal behavior based upon Guilford's model. They concluded that relatively little of the teacher-pupil interaction moved out of the lowest level of thinking, namely, "cognitive-memory." Sanders (1966) used a different category system for levels of thinking but reached basically the same conclusion.

In the area of supportive emotional climate, Medley and Mitzel (1959) utilized the 1958 Observation Schedule and Record (OScAR) to study the relation between teacher-pupil behavior and pupil growth in reading. Variations of OScAR were used by other researchers in looking at criticism and teacher control in relation to pupil growth in spelling, reading, and creativity (Soar, 1972).

Probably one of the most important systems developed was that of Flanders (1965). A major series of studies was based upon Flanders Interaction Analysis in conjunction with other systems. A comprehensive review of research based on Interaction Analysis cautiously reported greater growth in subject matter and more favorable attitudes on the part of pupils associated with more indirect teaching behavior, teacher flexibility and teacher sustained acceptance of pupil ideas.

Beginning with the work of Medley and Mitzel (1958, 1959) the use of multiple systems began to appear. Soar believes that their OScAR really was a combination of two observational processes, i.e., Withall's Technique (a forerunner of Flanders Interaction Analysis) combined with a sign system which involved groupings, subject matter, materials, and nonverbal communication. With the advent of the use of multiple systems the research became more complex. As more variables were added in terms of the process dimension, increases not only of non-significant results, but unexpected reversals and conflicting results were obtained.

Soar concludes:

> ". . . that the classroom is an infinitely complex social system, and that dimensions of it which seem subjectively to be aspects of the same complex may, indeed, be independent of each other, and may relate in very different ways to the same measures of pupil growth. . . The other side of the coin is that the same measure of classroom behavior will produce quite different relations with different measures of pupil growth. (p. 197)

However, in light of much of the research of the 1960's as here reviewed it seems that a classroom supportive of pupil growth will likely have a teacher who uses low criticism of pupils, is praising, accepts, and uses pupil ideas, and utilizes a minimum of control and restrictive direction (Soar, 1972).

The 1970's saw a multiplicity of research approaches to find the answer to the question of "What makes a good teacher?" The Process-Product approach includes those correlational studies which primarily look at the relationship between a teacher behavior variable or set of variables, and average student achievement scores. A review of studies concerning teacher behavior and pupil learning of basic skills in the elementary grades is provided by Rosenshine (1976), Medley (1977), and Good (1979). Brophy (1979) suggests that sufficient replications exist to provide dependable knowledge supportive of several generalizations, among which are:

1. Teachers make a difference (Good, Biddle, & Brophy, 1975; Rakow, Airasian, & Madaus, 1978).

2. No evidence appears to support the idea of generic teaching skills.

3. There appears to be supportive evidence for clusters of behaviors including: teacher expectations/role definitions, classroom management skills (Evertson & Anderson, 1978), and direct instruction (Good, 1979).

Large scale field correlational studies at the elementary level (such as Brophy & Evertson, 1976) appear to support evidence of these cluster behavior patterns. Also data concerning direct instruction seem to generalize to upper grades if the primary goal of instruction is basic skill mastery. (Brophy underlines a note of caution here.)

Peterson (1979) raises questions about direct instruction and cautions that, indeed, the effectiveness of this style of teaching may well depend on the type of student who is being taught.

Hunter (1979) looks upon decision making by the effective teacher

as a "synthesis of science and sensitivity to a situation" (p. 63). She thus equates professional preparation of the future teacher to that of the future doctor. It "requires the professional to learn, internalize, and implement the contributions of science to increased productive human functioning" (p. 62). Shavelson (1973) actually defines decision making as the basic teaching skill.

Powell (1979) reports new research evidence to support the significance of time on task. Citing results of the Beginning Teacher Evaluation Study (BTES) conducted in California, she emphasizes the importance of three major concepts: students should experience extensive content coverage, high success rate, and active involvement in their work. Her conclusion is that these concepts only serve to emphasize the complexity of teaching in today's classrooms. Fisher, Marliave, and Filby (1979), in discussing Academic Learning Time as a part of the California BTES (Fisher, et al., 1978), define a similar trio of ingredients, i.e., allocated time, engagement time, and success rate.

A dimension in teaching effectiveness which has further appeared to be of interest to researchers is that of teachers' instructional interpersonal relations styles. Hosford (1978) refers to the "silent curriculum," which is created as teachers teach. Aspy and Roebuck (1977) report positive and significant relationships between teachers' level of interpersonal functioning and student attendance, gain on achievement test scores, and enhanced self concept. Further, they report that these interpersonal relation skills, or facilitating behaviors, can be taught to teachers, and thus teachers can change as a result. Research into teacher personality style, or aspects of that style, may provide a basis for

developing strategies to improve the silent curriculum. Chu (1979) reports that several studies already point to the possibility that teacher sense of efficacy may well be a major element in determining what happens in the classroom, and that this sense of efficacy has an effect on student achievement and attitudes. Chu raises the question of whether it is possible to provide training and thereby bring about change in a teacher's sense of efficacy.

Similar questions might be asked about other aspects of teacher personality style. For example, Hosford and Schroder (1974) point to an intuitively perceived X-factor which seems to help identify "good" teachers. It seems to operate importantly in assessment of teacher performance. Can research data be gathered objectively to provide information on an elusive dimension which seems to be related to teacher effectiveness?

Much of the literature reviewed points toward a major premise that differences between truly effective and ineffective teachers are often associated with their interpersonal skills in dealing with the students, peers, administrators and parents. These skills are generally evaluated by supervisors and administrators close to the teacher. Such professional judgments have always operated in teacher evaluations, with the major problem being that of objectivity--that is, in demonstrating that such judgments are free of feeling associated with personal relationships between evaluators and those being evaluated.

RESEARCH MODEL

We present for consideration a model for research in teacher evaluation which deals with collecting objectively-gathered judgments which

may be quantitatively related to product and teacher effectiveness as defined in terms of both learner progress and university/school district assessment.

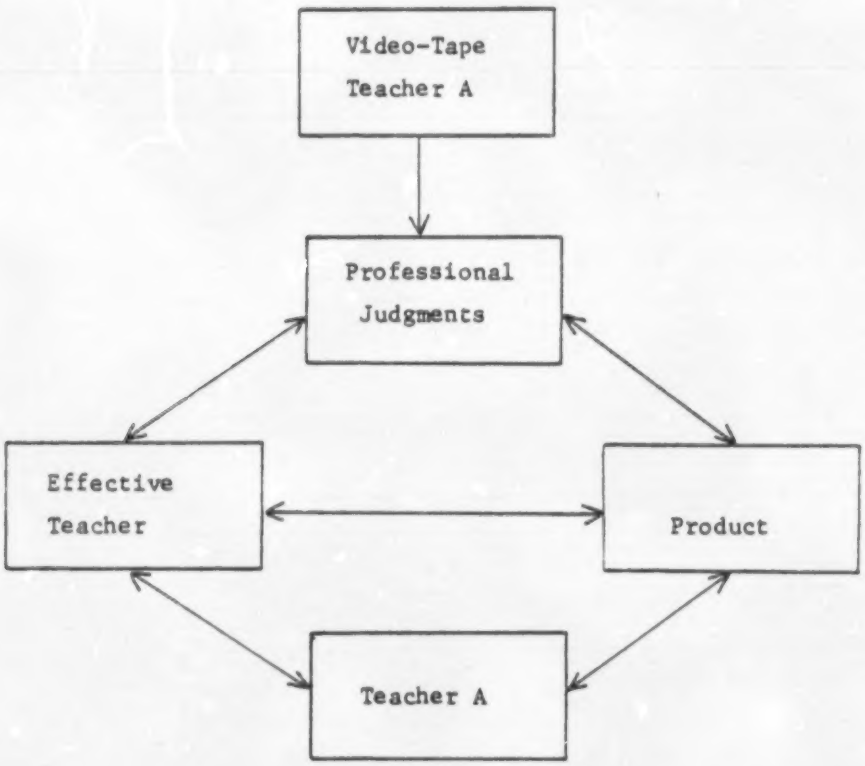

Figure 1.

Teacher Evaluation Model

To date, our research has dealt mainly with the connection of the Video-tape and Professional Judgments segments of the model and in a special, but limited sense, with the Effective Teaching and Product segments. We have been rigorous and consistent in our own definition of professional judgments. We have used only those gathered under the

following conditions:

1. Evaluators are experienced professional educators. (Student teaching or more.)

2. Four-minute observation in all cases.

3. The teachers being evaluated are not known to those who are doing the evaluation.

4. Evaluators respond to only one question with Yes, Maybe, No. The question is: Would you want this teacher next year for your child, or little brother or sister?

The development of this procedure for gathering professional judgments has evolved over the last eight years through a series of three studies.

1. Hosford-Luna Pilot Study (Hosford & Neuenfeldt, 1979). In 1973, two teachers, a graduate student and Hosford, were taped while teaching episodes in a fifth grade class. Each episode was from five to nine minutes in length. Two of the "teachers" taught two episodes--the second purposely poorer than the first. The "teachers" were numbered in the order in which they appeared on the tape. The identifying numbers were then placed on evaluation forms providing three categories to be checked. Children checked one of three categories: I want this teacher next year, I don't care, or I don't want this teacher. Similarly, the adult evaluators responded to the question, "Would you want this teacher for your own fifth grade child, brother or sister?" by checking a category of Yes, Maybe, or No. The evaluators thus indicated the category that best described their feelings toward the particular "teacher" in each episode.

All six of the taped recordings were viewed and evaluated by two

fifth grade classes in two schools; 16 beginning education students at New Mexico State University enrolled in Education 111; seven student teachers at New Mexico State University; and 31 professional educators enrolled in two graduate courses at NMSU (Education 468 and Education 624). Findings included:

A. Four minutes is the optimum viewer time.

B. Professional educators achieved internal consensus.

C. In some cases, the professional educator group consensus differed from that of fifth grade groups.

D. Low SES fifth grade group evaluations differed from those of the high SES fifth grade group.

E. The same teacher, by using different teaching styles, can get Yes and No responses.

2. Hosford-Schroder Pilot Study (1974). In Fall Semester, 1972, all students enrolled in Education 211 participated in microteaching sessions in which they demonstrated the skill, "Obtaining Attending Behavior." One hundred eight students were videotaped during these microteaching sessions. Each "teacher" taught three, four, or five peers for a period of two to seven minutes. Fall Semester, 1973, a 40 minute section of these videotapes containing presentations by 13 of the Education 211 students was selected. This 40 minute tape was shown to six groups. When shown to a graduate class (Education 535), the evaluators were asked to answer with one of three responses the following question for each prospective teacher viewed, "What are the chances for this person to become an effective teacher?": Good (looks very strong), Marginal (could go either way, exhibits strengths and weaknesses about equally), Poor (looks very weak).

The Education 5[illegible] class was made up of 12 full-time master's level students and 13 full-time public school classroom teachers.

Three doctoral-level students and two professors followed the same viewing and evaluation processes. All of these viewers were experienced teachers: three had supervisory experience with student teachers. The same 40 minute videotape was shown to a junior high school ninth grade civics class (N=24) and to a senior high school sociology class (N=21).

Five ninth grade students, recommended by the principal as representative of the ninth graders in his school, were employed at minimum wage. to stay after school one hour a day, two days a week, until they had viewed and evaluated all 108 videotapes. Students were asked to respond in one of three categories, Yes, Maybe, No, to the question, "If you had your choice, would you want to be in this teacher's class?" When an elementary school lesson was presented by a prospective teacher on the videotape, the evaluators were told to consider the question from a younger person's point of view such as a brother or sister, or "Do you think this person would be a good teacher?"

Ratings for each prospective teacher were calculated for each group of evaluators. Good or Yes = +1; Marginal or Maybe = 0; and Poor or No = -1. The 13 "teachers" were then ranked according to the ratings.

Findings:

A. Four minute optimum viewer time was confirmed.

B. All groups achieved internal consensus.

C. The five paid students spoke effectively for all secondary students (correlations of ratings significant at .05 level).

D. Low, positive correlations were found between consensus of five paid students and all experienced educator groups.

E. Consensus became more definite as group grade level of education increased.

Because of the low correlations between student and professional educator ratings, and the overall consistent agreement of all professional educator groups with other professional educator groups, a decision was made to limit judgments to those with teaching experience, including those who had completed student teaching. A follow-up study is needed since the reliability of group responses and the high consensus level achieved might be due to age alone.

3. Hosford-Neuenfeldt Study (Neuenfeldt, 1977; Hosford & Neuenfeldt, 1979). As a result of the two pilot studies, valuable questions were derived which influenced the design of a third study. This study was conducted in two stages. In Stage One, five graduate students in the Elementary and Secondary Education Department at New Mexico State University taught two, four-minute segments to fifth grade classes in Las Cruces, New Mexico. One segment was taped of each graduate student's normal teaching behavior (normal: defined by the graduate students upon viewing their tapes and judging them to be representative of their teaching behavior). The second segment was taped while each graduate student demonstrated experimental teaching behavior (experimental: defined as when the teacher knowingly and consistently attempts to violate one or two desirable behaviors). The taped behaviors were then edited onto two tapes, each tape containing only one of each teacher's teaching behaviors. The two tapes, A and B, were viewed and evaluated by more than 300 raters including groups of teachers, student teachers, and public school students. The results of Stage One were then used in conjunction with a field test of the procedure.

In Stage Two, the field test conducted in the Menomonie, Wisconsin, area, videotapes were made of eight junior and senior high school teachers in one of their regular classes. Videotapes also were made of six college mathematics teachers each teaching a regular class. The teachers were given the option of having as many minutes taped as they desired. They all chose to use the basic segment, as it was first taped. Rater groups were again experienced teachers, education students and public school students. The questionnaire form used provided only three options of response: Yes, Maybe, and No.

As a significant dimension of this study, ratings were also obtained from supervisors of the teachers independent of the videotape procedures. These ratings, which were obtained through the regular processes used by each school organization gathered throughout the entire school year, were then compared with those derived from the videotape evaluations.

Findings:

A. Professional educators achieve a higher Yes-No consensus in their evaluation of videotaped teaching segments than do other groups. Significant agreement is achieved by all groups when they view and evaluate segments that prompt definite Yes or No reactions from professional educator groups.

B. When a strong Yes or No consensus is achieved by professional educators, viewer variables of sex, age and grade level of teaching experience plays no significant part.

C. Student groups do not produce tally distributions comparable to professional groups. As the maturity of student evaluators increases, their tally distributions become more like those of the professional groups.

D. Consensus is less frequently achieved by all viewer groups when videotape segments fail to prompt mostly Yes or No reactions from the professional groups.

E. Supervisory evaluations agreed with the previously-obtained videotape evaluations. In all cases, relatively high positive correlations were obtained. However, some cases were not statistically conclusive and additional research is needed relating videotape and existent supervisory evaluations before a stronger assertion can be made.

In summary, the objectively-gained judgments obtained through videotape procedures described in these studies apparently correlate with judgments being made in the field. These same judgments correlate most highly when they relate to videotape segments that commonly prompt definite Yes or No evaluations. Professional educators all obtain high intergroup agreement in their evaluations of the videotape segments. Viewer age, sex and grade level of teaching experience do not appear to be significant variables (Hosford & Neuenfeldt, p. 421).

Questions still remain:

A. Will preservice education students identified as definite Yes or No types later confirm these evaluations in their overall college performances and especially in their student teaching performance?

B. Will videotape diagnosis prove more predictive than other variables such as competency tests in cognicive areas, ACT scores, overall grade point average, or other available predictors?

C. Will the variables of ethnicity or physical traits on the part of viewer-viewees significantly affect videotape evaluations?

4. Hosford-Martin Study I (1980). In an attempt to answer some

of these questions, Hosford and Martin began a new study, funded through a mini-grant from the Educational Research Center at New Mexico State University. This study is still in progress. During the Spring semester, 1978, all Education 211 elementary education majors were videotaped during their first microteaching experience. Seventy-six students received the same preparation for this assignment, and these students were then videotaped under the supervision of one instructor using the same directions and facilities. Seventy-one students signed permission slips authorizing the use of their tapes for research purposes.

In Summer 1978, six experienced teachers were employed to view all 71 videotape segments and were directed to respond only to the question: "Would you want this teacher for your child or little brother or sister next year?" They recorded their responses on a form which provided three options: Yes, Maybe, and No. These six teachers were representative of K-12 teaching levels as well as two through nine years of classroom experience. There was also representation in terms of ethnicity, sex, and age. They viewed no more than 11 tape segments on any one day. Each teaching segment viewed was exactly four minutes long.

Four steps were followed in testing the hypothesis that these six teachers were representative of any group of professional educators who might view the taped segments.

Step 1. The 71 tape segments were rank-ordered on the basis of the votes of the six teachers. Six of these segments received a unanimous vote of Yes while only three received a unanimous vote of No. A new tape (labeled Tape 8) was edited to contain the three unanimous No segments, three segments from the Maybe group, and four segments conveniently

selected from the unanimous Yes group. These ten segments were then mixed in order of appearance on Tape 8.

In Summer 1978, 31 teachers in a graduate course (Education 535) viewed Tape 8 and evaluated all 10 segments with the response of Yes, Maybe, or No to the question, "Would you want this teacher for your child or little brother or sister next year?"

The responses of these 31 teachers were compared with those of the original six teachers hereafter called the 0 - 6 group. A Pearson product moment correlation coefficient of .93 was obtained. The critical value was .765 at the .005 level of confidence.

Step 2. In Fall 1978, the first three segments of Tape 8, containing a Yes, Maybe, and No segment as previously assessed by the 0- 6 group, were shown to 217 teachers and administrators in the El Paso schools. Of the 217 teachers, 119 taught grades 4-8. The responses of these 119 teachers tested against the 0 - 6 group responses yielded a correlation of .99. The 98 remaining were teachers of grades 9-12. Their responses, tested against the 0 - 6 group, yielded a .99 correlation. Because of these high correlations in Step 2, a third step was undertaken to support the hypothesis that any six professional educators would be sufficient to speak for larger groups.

Step 3. The responses obtained from the 31 teachers in Step 1 were used for this purpose. Using a table of random numbers, six of the 31 were selected. Following the same procedure, four other subsets of six were selected. Thus, five subsets, each containing six teacher-responses, were obtained. The mean response of each subset was tallied and tested against the 0 - 6 group mean response. For each of the five subsets, the

corresponding correlations with the 0 - 6 group were .87, .94, .89, .95, and .91 respectively.

Step 4. A year later, Fall 1977, Tape 8 again was shown to 13 professional educators. Their responses were tallied and compared with both the 0 - 6 group and the group of 31. Correlations obtained respectively were .79 and .85. These findings point to a high reliability of inter-group means. To date, our findings indicate that:

A. The teaching experience grade level of the viewing professional educator makes no difference.

B. Viewer consensus is obtained from professional educators viewing four minute videotape segments of teaching behavior.

C. Any six professional educators will provide an adequate sample for obtaining general consensus.

D. Professional educator consensus will be reliably obtained from one year to the next.

Some obvious limitations should be noted:

A. The study only shows judgments confirmed by judgments.

B. Data have not yet been obtained to check predicted success (indicated by Yes and No votes) against performance in student teaching.

C. The viewer variables of age, sex, ethnicity, and years of teaching experience have not yet been examined in this study.

5. Morrissey Study. Currently in progress is a doctoral dissertation undertaken by Bill Morrissey which has as one of the components a follow-up on the Spring 1978, Education 211 elementary major students. Utilizing data obtained from the previous Hosford-Martin studies, Morrissey is concerned with students who were enrolled in student teaching

in elementary schools in Fall 1979. Within this group are a number of subjects from the Spring 1978, Education 211 videotaped microteaching group.

He has collected: (A) 1978 Education 211 videotape evaluations compared with pre-student teaching videotape evaluations; (B) Education 211 videotape judgment compared with post-student teaching field evaluations; (C) comparisons of the Education 211 videotape judgments with ratings by student teaching supervisory personnel, i.e., professor, field supervisor and cooperating teacher; and (D) Education 211 videotape judgment compared with other predictors, ACT, GPA, etc.

6. Hosford-Martin Study II. Also in progress is a replication study of the previous Hosford-Martin work with the Spring 1978, Education 211 elementary education major students. In Fall 1979, videotapes were made of Education 211 secondary education major students. Permission slips authorizing the use of these taped microteaching segments for research purposes have been obtained. Editing of these tapes is currently underway.

REVIEW OF CURRENT RELATED RESEARCH ON THE USE OF VIDEOTAPE RECORDINGS

Despite extensive research in teacher effectiveness, it is only in the last ten years that a few findings have been validated. With recent developments in the area of accountability there is a need to continue to pursue this elusive topic. A reliable evaluation procedure of teacher performance would be a significant move in this direction. The use of videotapes and evaluation of the videotape segments may provide such a

procedure. However, current research in this specific area appears to be very limited. Many evaluation procedures exist using videotaping methods, but these generally are in the area of student teacher evaluations (Neuenfeldt, 1977). Goldberg and Mayerberg (1973) used videotapes to test student ability to evaluate differentially various types of nonverbal teacher behavior. Hodge (1975) would appear to believe in the value of the use of videotapes as a means of excluding observer bias for recording and evaluating nonverbal teacher behavior.

In a very recent literature search on the use of videotapes in teacher evaluation, over 100 entries were produced. Of these, 21 were found in journals, 36 were in Dissertation Abstracts, and 55 were located in ERIC. Excluding Hosford publications, only four of the 21 journal articles appeared to be related directly to our area of research. Three of the 36 dissertations and three of the 55 ERIC articles were relevant.

Several other studies focused on the use of videotapes in conjunction with, or in support of, various assessment scales, ratings and inventories. These have met with apparently varying degrees of success (James, 1978; Love, et al., 1978; Cohen & Bengston, 1975; Kravas, 1975; Wilson, 1974).

Use of videotapes in evaluating preservice teacher performance or as a tool to improve teacher classroom effectiveness is reported by De Genaro (1969), Mears (1977), Smith (1969), and Wenner (1969). These studies primarily emphasized the use of videotapes for supervisor feedback or for self-appraisal.

The Beginning Teacher Evaluation Study (BTES), Phase II, was a research project on effective teacher behavior which was based on videotaped recordings (Sandoval, 1976). It is a significant report for all

those who are interested in this particular area of teacher effectiveness research. Findings in this report support evidence that: (1) teacher behavior on tape appeared to be consistent with teaching style over a period of time; (2) videotape recordings of pupils were problematic and therefore the camera should focus fairly closely on the teacher with few children visible; and (3) categories should be limited when obtaining ratings of the tapes to avoid obfuscation of data.

An interesting aspect of the BTES study involved identifying the five teachers with greatest pupil achievement gains and the five teachers with the least pupil achievement gains at both fifth and second grade levels in reading and math classes. They speculated that, perhaps, the top or bottom five teachers may be truly different from their colleagues and may tend to break the usual rules of good or poor teaching.

SUMMARY

A very small but increasingly important body of literature relating the use of videotape recordings to evaluation of teacher effectiveness seems to be emerging. Much research needs yet to be done. We propose the following questions for testing further segments of our Teacher Evaluation Model:

1. Can videotapes of both extremes (Yes and No ratings) be used effectively as models in teacher training?

2. Can a study of such extremes lead to identification of effective and ineffective characteristics, behaviors, or process-product variables associated with the teacher evaluation models previously described?

3. Will videotape predictions of the extremes be confirmed by

district assessments during the teachers' first or second year teaching in the field?

4. Can professional judgments currently made in the field identifying extreme cases be predicted through the four minute videotape evaluations?

5. Do either or both field judgments (3 and 4) correlate with product (standardized achievement test results) or with other measurements of product (silent curriculum elements)?

6. Are professional educators the only reliable, valid viewer groups?

7. Why does SES of student viewers affect evaluations? To what degree?

Results obtained thus far have been encouraging. The use of videotapes for early identification of pre-service students with potential problems and of those who are clearly superior would have significant ramifications for teacher education programs. The extremes depicted on these videotapes could serve as useful models for students in teacher training programs. In addition, these same tapes could provide material for further study of what constitutes desirable and undesirable teacher behaviors.

We present here our model for teacher evaluation in the hope that a valid and reliable procedure for predicting teacher effectiveness may ultimately be developed.

REFERENCES

Amidon, E., & Hunter, E. Improving teaching: The analysis of classroom verbal interaction. New York: Holt, Rinehart & Winston, 1966.

Aspy, D., & Roebuck, F. Kids don't learn from people they don't like. Amherst, Mass.: Human Resource Development Press, 1977.

Brophy, J. E. Teacher behavior and student learning. Educational Leadership, 1979, 37, 33-38.

Brophy, J., & Evertson, C. Learning from teaching: A developmental perspective. Boston: Allyn and Bacon, 1976.

Brophy, J. E., & Good, T. L. Teacher-student relationships. New York: Holt, Rinehart & Winston, 1974.

Burkhard, M. I. Discernment of teacher characteristics by T.A.T. sequence analysis. Journal of Educational Psychology, 1962, 53, 279-287.

Chu, L. Toward a humanistic behaviorism: Self efficacy in multicultural education. Dialogue Series, New Mexico State University, Las Cruces, October 23, 1979.

Clark, C. M. Five faces of research on teaching. Educational Leadership, 1979, 37, 29-32.

Cohen, S. J., & Bengston, J. K. A psychophysical investigation of factors affecting teacher-observers' judgment. Las Cruces: ERIC, 1975. (ERIC Document Reproduction Service No. ED 104 942)

DeGenaro, A. P. Experimental use of the video tape recorder as an evaluative instrument and observational tool in the supervision of student teachers of physical education (Doctoral dissertation, The Ohio State University, 1969). Dissertation Abstracts International, 1969, 30/07A, 2879.

Evertson, C., & Anderson, L. The classroom organization study: Interim progress report. Report No. 6002. Austin: Research and Development Center for Teacher Education, The University of Texas at Austin, 1978.

Fisher, C., Marliave, R., & Filby, N. Improving teaching by increasing academic learning time. Educational Leadership, 1979, 37, 52-54.

Fisher, C. W., Filby, N., Marliave, R. S., Cohen, L. S., Dishaw, N. M., Moore, J. E., & Berliner, D. C. Teaching behaviors, academic learning time and student achievement. Final report of Phase III-B, Beginning Teacher Evaluation Study. San Francisco: Technical Report U-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1978.

Flanders, N. Analyzing teaching behavior. Reading, Mass.: Addision-Wesley, 1970.

Flanders, N. Teacher influence, pupil attitudes, and achievement. Cooperative Research Monograph No. 12, OE-25040. Washington, D.C.: United States Department of Health, Education and Welfare, Office of Education, 1965.

Gallagher, J., & Aschner, M. J. A preliminary report: Analysis of classroom interaction. Merrill-Palmer Quarterly, 1963, 9, 183-194.

Goldberg, G., & Mayerberg, C. K. Emotional reactions of students to nonverbal teaching behavior. Journal of Experimental Education, 1973. 42, 29-32.

Good, T. Teacher effectiveness in the elementary school: What we know about it now. Journal of Teacher Education, 1979, 30, 52-64.

Good, T., Biddle, B., & Brophy, J. Teachers make a difference. New York: Holt, Rinehart & Winston, 1975.

Hodge, R. L. Recording nonverbal behavior for effective teacher evaluation. Contemporary Education, 1975, 46, 189-193.

Hosford, P. L. An instructional theory. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973.

Hosford, P. L. Your choice of teachers for your son and daughter next year? New Mexico School Review, 1973, 51, 20-22.

Hosford, P. L. The silent curriculum: Its impact on teaching the basics. Educational Leadership, 1978, 36, 211-215.

Hosford, P. L., & Martin, J. V. Prospective teacher evaluation via videotape. Action In Teacher Education, 1980, in press.

Hosford, P. L., & Neuenfeldt, J. C. Teacher evaluation via videotapes: Hope or heresy? Educational Leadership, 1979, 36, 418-422.

Hosford, P. L., & Schroder, A. Can the ex-factor of teaching be used in research? Las Cruces: New Mexico State University, 1974. (ERIC Document Reproduction Service No. ED 100 920)

Hunter, M. Teaching is decision making. Educational Leadership, 1979, 37, 62-64, 67.

James, B. N. An examination of an approach for using pupil outcome scores to validate assessments of teacher behaviors, 1978. (ERIC Document Reproduction Service No. ED 161 944)

Kravas, C. H. Using videotape vignettes to measure teacher affective sensitivity. Educational Technology, December 1975, 15, 13-16.

Love, J., et al. The generalizability and concurrent validity of an instrument to describe college teaching, 1978. (ERIC Document Reproduction Service No. ED 167 611)

Means, R. K. Instrument for accountability: The videotape. Health Education, March-April, 1977, 8, 7.

Medley, D. M. A summary of teacher effectiveness research. Charlottsville, Virginia: Department of Research Methods, University of Virginia, 1977.

Medley, D. M. Teacher competence and teacher effectiveness: A review of process-product research. Washington, D.C.: American Association of Colleges for Teacher Education, 1977.

Medley, D. M., & Mitzel, H. E. Some behavioral correlates of teacher effectiveness. Journal of Educational Psychology, 1959, 50, 239-246.

Medley, D. M., & Mitzel, H. E. A technique for measuring classroom behavior. Journal of Educational Psychology, 1958, 49, 86-92.

Neuenfeldt, J. C. An investigation of an alternative method of evaluating classroom teaching (Doctoral dissertation, New Mexico State University, 1977). Dissertation Abstracts International, 1978, 38, 7121A-122A.

Peterson, P. L. Direct instruction: Effective for what and for whom? Educational Leadership, 1979, 37, 46-48.

Powell, M. New evidence of old truths. Educational Leadership, 1979, 37, 49-51.

Rakow, E., Airasian, P., & Madaus, G. Assessing school and program effectiveness: Estimating teacher level effects. Journal of Educational Measurement, 1978, 15, 15-21.

Rosenshine, B. Recent research on teaching behaviors and student achievement. Journal of Teacher Education, 1976, 27, 61-64.

Rosenshine, B., & Berliner, D. Academic engaged time. British Journal of Teacher Education, 1978, 4, 3-16.

Ryans, D. G. Characteristics of teachers, their description, comparisons and appraisal: A research study. Washington, D.C.: American Council on Education, 1960.

Sanders, N. M. Classroom questions: What kind? New York: Harper and Row Publishers, 1966.

Sandoval, J. The evaluation of teacher behavior through observation of videotape recordings. Beginning teacher evaluation study: Phase II, 1973-74, Final Report: Volume III.3, 1976. (ERIC Document Reproduction Service No. ED 127 368)

Shavelson, R. J. What is the basic teaching skill? Journal of Teacher Education, 1973, 24, 144-151.

Smith, R. H. A comparison of instructional evaluations by teachers and supervisors using video-tape (Doctoral dissertation, University of Minnesota, 1969). Dissertation Abstracts International, 1969, 30/04A, 1384.

Soar, R. S. Teacher-pupil interaction. In J. R. Squire (Ed.), A new look at progressive education. Washington, D.C.: Association for Supervision and Curriculum Development, 1972.

Wenner, H. M. An assessment of video tape as a tool in the improvement of teacher classroom effectiveness (Doctoral dissertation, University of Minnesota, 1969). Dissertation Abstracts International, 1969, 31/03A, 941.

Wilson, L. Assessing teacher skills: Necessary component of individualization. Phi Delta Kappan, 1974, 56, 207-209.